Theoretical Computer Science 99 (1992) 327–334
Elsevier

Note

Puissances de mots et reconnaissabilité des points fixes d'une substitution*

Brigitte Mossé
Université de Provence, 3, Place Victor Hugo, 13331 Marseille Cedex 3, France

Communicated by D. Perrin
Received June 1990
Revised January 1991

Résumé

Mossé, B., Puissances de mots et reconnaissabilité des points fixes d'une substitution, Theoretical Computer Science 99 (1992) 327–334.

Soit u un mot infini non périodique point fixe d'une substitution primitive σ; nous montrons que pour N assez grand u ne contient aucune puissance N-ème de mot. Nous donnons ensuite une condition nécessaire et suffisante pour que σ soit reconnaissable au sens de Host et Queffélec, et nous montrons qu'il suffit de "bilatéraliser" cette définition pour obtenir une propriété vérifiée cette fois par toutes les substitutions primitives à point fixe non périodique.

1. Introduction et notations

C'est en cherchant à construire des mots ne contenant aucun cube que Thue [8] a été conduit au début du siècle à considérer une suite point fixe de la substitution définie par $\sigma(a)=ab$ et $\sigma(b)=ba$. Cette suite, introduite en fait par Prouhet en 1851, fut redécouverte en 1921 par Morse qui lui donna son nom. Depuis Karhumäki a montré que le mot de Fibonacci, point fixe de la substitution définie par $\sigma(a)=ab$ et $\sigma(b)=a$ ne contient pas de puissance quatrième de mot (cf. [6]).

Dans la Section 2, nous montrons plus généralement que, dès que u, point fixe d'une substitution primitive, n'est pas périodique, il existe un entier naturel N tel qu'aucune

* Travail réalisé dans le cadre du PRC "Mathématique Informatique".

puissance N-ème de mot n'apparaisse dans u. Bien que ne fournissant pas le plus grand exposant effectif, la démonstration du Théorème 2.4 permet d'en calculer aisément un majorant.

On entend par reconnaissabilité pour une substitution σ une notion exprimant qu'on peut retrouver la structure d'un facteur d'un point fixe de σ dès qu'il est assez long. C'est une hypothèse qu'on est naturellement amené à faire lorsqu'on étudie le système dynamique associé à ce point fixe.

La première formalisation de cette idée est due à Martin [4], mais il énonce à ce sujet un théorème peu convaincant. Host et Queffélec donnent alors une définition moins exigeante et introduisent le terme "reconnaissable" [2, 5].

Host affirme dans [2] qu'une substitution à point fixe non périodique non postfixée est reconnaissable. De plus Rauzy a donné une démonstration courte du fait qu'une substitution de longueur constante injective sur les lettres et à point fixe non périodique est reconnaissable ([5]).

Le Théorème 3.1 donne une caractérisation des substitutions primitives reconnaissables; toutes ne le sont pas. Le Théorème 3.1bis montre que la notion bilatéralisée de reconnaissabilité, sans inconvénient pour l'étude du système dynamique associé, est elle vérifiée par toutes les substitutions considérées.

De plus, les problèmes de la reconnaissabilité de σ, de la périodicité de u, et de la présence de puissances de mots dans u sont extrêmement liés: par exemple, la périodicité de u est incompatible avec le fait qu'une infinité d'itérés de σ soient reconnaissables; si v est un élément de $\mathcal{A}^*$ et si u n'est pas périodique, v^N ne peut pas apparaître dans u pour tout N, par minimalité (c'est faux en général dans le cas non primitif); il est naturel enfin, pour étudier les puissances de mots intervenant dans u, d'utiliser l'éventuelle reconnaissabilité de σ pour faire décroître la longueur des mots considérés. Dans la Section 3, la démonstration des Théorèmes 3.1 et 3.1bis précisera ce lien.

Soit $\mathcal{A}$ un ensemble fini de cardinal $g \geqslant 2$, appelé alphabet; les éléments $\alpha_1, \ldots, \alpha_g$ de $\mathcal{A}$ sont appelés lettres. On note $\mathcal{A}^*$ l'ensemble des mots non vides de longueur finie sur $\mathcal{A}$. Si $v \in \mathcal{A}^*$, $|v|$ note sa longueur, et si $x \in \mathcal{A}^{\mathbb{N}}$, $x[i, j]$ note le mot $x_i x_{i+1} \ldots x_j$ $(j \geqslant i)$.

Nous désignerons par σ une substitution sur $\mathcal{A}$, c'est-à-dire une application de $\mathcal{A}$ dans $\mathcal{A}^*$ prolongée par concaténation à $\mathcal{A}^*$ et à $\mathcal{A}^{\mathbb{N}}$, et admettant un point fixe u (cf. [1]).

Nous supposerons de plus que σ est primitive, c'est-à-dire qu'il existe un entier naturel n tel que, pour tout couple (a, b) de lettres, b apparaît dans le mot $\sigma^n(a)$. Ceci revient à dire que la matrice associée à σ, définie par

$$M = (M_{ij})_{1 \leqslant i, j \leqslant g}, \text{ où } M_{ij} \text{ est le nombre d'apparitions de } \alpha_i \text{ dans } \sigma(\alpha_j),$$

admet une puissance strictement positive.

Soit T le décalage défini sur $\mathcal{A}^{\mathbb{N}}$ par

$$T((x_n)_{n \in \mathbb{N}}) = (x_{n+1})_{n \in \mathbb{N}}.$$

Si l'on note K_u la fermeture de l'orbite de u sous l'action de T pour la topologie produit sur $\mathscr{A}^{\mathbb{N}}$, on sait que le système dynamique (K_u, T) est strictement ergodique. En particulier, ce système est minimal: tout élément de $\mathscr{A}^*$ apparaissant dans u y apparaît une infinité de fois, avec des lacunes bornées (cf. [1]).

Nous noterons L_u l'ensemble des facteurs du mot infini u.

Notons maintenant

$$E_k = \{0\} \cup \{|\sigma^k(u[0, p-1])|;\ p > 0\}\ .$$

Définition 1.1. On dit que σ est reconnaissable s'il existe un entier $L > 0$ tel que si $u[i, i+L-1] = u[j, j+L-1]$, et $i \in E_1$, alors $j \in E_1$.

Nous introduisons la notion voisine suivante, la reconnaissabilité se confondant avec une notion de "reconnaissabilité à droite":

Définition 1.2. On dit que σ est bilatéralement reconnaissable s'il existe un entier $L > 0$ tel que si $u[i-L, i+L] = u[j-L, j+L]$, et $i \in E_1$, alors $j \in E_1$.

2. Puissances de mots dans un point fixe de substitution

Définition 2.1. Un mot infini u est dit ultimement périodique s'il existe des mots v_1 et v de longueurs finies tels que: $u = v_1 v^{\omega}$; il est dit périodique (ou purement périodique) dans le cas où v_1 est un suffixe de v.

Remarquons que dans le cas où u est point fixe d'une substitution primitive, l'ultime périodicité entraine la périodicité.

La notion de mot primitif (sans rapport avec l'hypothèse σ primitive), utilisée dans [7] pour étudier la périodicité des points fixes de morphismes, est définie ainsi (cf. [3]):

Définition 2.2. Un mot v de longueur finie est dit primitif s'il n'est pas puissance d'un autre mot que lui même.

Propriété 2.3. *Soit $w = v^n$, où v est un mot primitif et $n \geqslant 2$. Si vw_0v est un facteur de w, alors w_0 est une puissance de v.*

Démonstration. Si w_0 n'est pas une puissance de v, il existe deux mots non vides v_1 et v_2 tels que $v = v_1 v_2 = v_2 v_1$, et on sait alors que v_1 et v_2 sont puissance d'un même mot primitif, d'où la propriété énoncée. □

Théorème 2.4. *Soit u point fixe d'une substitution primitive, u non périodique. Il existe un entier N tel que dans u n'apparaisse pas de puissance N-ème de mot.*

Lemme 2.5. *Soit u point fixe d'une substitution primitive. S'il existe un mot primitif v et deux entiers naturels N et p tels que:*

(1) *Pour tous éléments a et b de $\mathscr{A}$, si ab appartient à L_u, $\sigma^p(ab)$ est un facteur de v^N,*

(2) $2|v| \leqslant \min\{|\sigma^p(a)|;\ a \in \mathscr{A}\}$,
alors u est périodique.

Démonstration. Pour tout élément i de $\{1, \ldots, g\}$, il existe d'après (1) et (2) un entier $n_i \geqslant 1$, un préfixe strict w_i de v et un suffixe strict v_i de v (c'est à dire de longueurs $<|v|$) tels que $\sigma^p(\alpha_i) = v_i v^{n_i} w_i$.

Si $\alpha_i \alpha_j$ est un élément de L_u, alors $v^{n_i} w_i v_j v^{n_j}$ est aussi dans L_u; v étant primitif, $w_i v_j$ est donc égal soit au mot vide, soit à v.

On en déduit que $u = \sigma^p(u)$ est périodique.

Démonstration du Théorème 2.4. Nous désignerons par v le plus petit entier tel que dans tout élément de longueur v de L_u apparaissent tous les mots de longueur 2 de L_u; l'existence de v est assurée par la primitivité de σ.

Soit l un entier naturel, $l > \frac{1}{2}\min\{|\sigma(a)|;\ a \in \mathscr{A}\}$.

Soit v un mot primitif de longueur l, et p le plus petit entier tel que:

$$2l \leqslant \min\{|\sigma^p(a)|;\ a \in \mathscr{A}\}.$$

Dans tout mot de longueur $(2\max\{|\sigma^p(w)|;\ w \in L_u,\ |w| = v\})$, apparaissent tous les $\sigma^p(ab)$, où ab appartient à L_u.

Si v^M appartient à L_u, on doit avoir d'après le lemme:

$$|v^M| = Ml < 2\max\{|\sigma^p(w)|;\ w \in L_u,\ |w| = v\},$$

et donc

$$M < 2\max\{|\sigma^p(w)|;\ w \in L_u,\ |w| = v\}\left(\tfrac{1}{2}\min\{|\sigma^{p-1}(a)|;\ a \in \mathscr{A}\}\right)^{-1}.$$

Notons N_l le membre de droite de cette inégalité; il n'y a pas dans u de puissance $(E(N_l)+1)$-ème de mot primitif de longueur l (E partie entière).

Or, pour tout élément a de $\mathscr{A}$, $|\sigma^p(a)|$ est de la forme $C_a\lambda^k + o(\lambda^k)$, où C_a est une constante strictement positive et λ la valeur propre dominante de la matrice associée à σ (cf. [1]). Ainsi N_l admet une limite, lorsque l tend vers l'infini, et les mots primitifs de longueur l, pour $l > \frac{1}{2}\min\{|\sigma(a)|;\ a \in \mathscr{A}\}$, n'apparaissent pas dans u au delà d'un certain exposant.

Il en est de même des mots de longueur $< \min\{|\sigma(a)|;\ a \in \mathscr{A}\}$, par minimalité, puisqu'ils sont en nombre fini.

On conclut finalement qu'il existe un entier N tel que dans u ne peut pas apparaître de puissance N-ème de mot, ce qui achève la démonstration du Théorème 2.4. $\square$

3. Théorèmes de reconnaisabilité

Introduisons pour la commodité de l'exposé les définitions suivantes:

Soit B un élément de L_u, et i et j deux entiers naturels tels que $B = u[i, i+|B|-1] = u[j, j+|B|-1]$. Nous dirons que B admet le même k-découpage au rang i et au rang j si $E_k \cap \{i, \ldots, i+|B|-1\}$ et $E_k \cap \{j, \ldots, j+|B|-1\}$ sont des translatés l'un de l'autre dans la translation de $j-i$ $(i \leqslant j)$.

- Soient $x_1, x_2, \ldots, x_{r+2}$ des éléments de $\mathscr{A}$, p un entier naturel, S un suffixe de $\sigma^p(x_1)$, P un préfixe de $\sigma^p(x_{r+2})$. Nous dirons que $[S, \sigma^p(x_2), \ldots, \sigma^p(x_{r+1}), P]$ est un p-découpage naturel du mot $B = u[i, i+|B|-1] = S\sigma^p(x_2)\ldots\sigma^p(x_{r+1})P$ au rang i s'il existe un entier naturel i' tel que $x_1x_2\ldots x_{r+2} = u[i', i'+r+1]$ et le bloc B admet le même k-découpage au rang i et au rang $|\sigma^p(u[0, i'-1])| + |\sigma^p(x_1)| - |S|$, pour tout $k \leqslant p$.

Considérons par exemple le mot de Morse u, point fixe de la substitution définie par $\sigma(a) = ab$ et $\sigma(b) = ba$, $u = abbabaabbaababbabaababbaabbabaab\ldots$.

Le mot bab n'a pas le même 1-découpage au rang 2 (où il admet le 1-découpage naturel $[ba, b]$) et au rang 11 (où il admet le 1-découpage naturel $[b, ab]$).

Théorème 3.1. *Soit σ une substitution primitive admettant un point fixe non périodique u. Pour que σ ne soit pas reconnaissable, il faut et il suffit que pour tout entier L, il existe n mot B de longueur L et deux éléments a et b de $\mathscr{A}$ tels que:*

le mot $\sigma(b)$ est un suffixe strict de $\sigma(a)$,

- *les mots $\sigma(a)B$ et $\sigma(b)B$ apparaissent dans u avec le même 1-découpage de B.*

Théorème 3.1bis. *Soit σ une substitution primitive admettant un point fixe non périodique u; la substitution σ est bilatéralement reconnaissable.*

Démonstration du Théorème 3.1. (1) Remarque préliminaire: Les éléments de E_p sont en quelque sorte synchronisants pour les problèmes de k-découpages. En effet, pour tout élément a de $\mathscr{A}$, la suite dont le n-ème terme est la première lettre de $\sigma^n(a)$ est ultimement périodique. Il existe ainsi un entier q (zéro convient si pour toute lettre a cette suite est purement périodique) tel que si i et j sont deux éléments de E_p $(p > q)$ vérifiant $u_i = u_j$, il existe un élément a de $\mathscr{A}$ tel qu'au rang i et au rang j apparaisse le mot $\sigma^{p-q}(a)$, cette expression donnant son $(p-q)$-découpage naturel aux rangs i et j.

La même remarque vaut à gauche des rangs appartenant à E_p, en considérant la dernière lettre des itérés d'une lettre donnée, et dans la limite du nombre de termes de u placés à gauche des rangs envisagés.

(2) Supposons que σ n'est pas reconnaissable. Soit k un entier naturel non nul fixé. Par non reconnaissabilité, il existe pour tout entier p un mot R_p et deux entiers naturels i et j, $i \in E_1$, $j \notin E_1$, tels que:

- $R_p = u[i, i+|R_p|-1] = A_p\sigma^p(a_0)\ldots\sigma^p(a_{k-1})$, où $a_{-1}, a_0, \ldots, a_{k-1}$ sont des lettres, A_p un suffixe de $\sigma^p(a_{-1})$ et $[A_p, \sigma^p(a_0), \ldots, \sigma^p(a_{k-1})]$ est un p-découpage naturel de R_p au rang i;
- $R_p = u[j, j+|R_p|-1] = A_pB_p\sigma^p(b_0)\ldots\sigma^p(b_{h-1})B'_p$, où $b_0, \ldots, b_{h-1}$ sont des lettres, B_p et B'_p des mots, B_p étant non vide, et $[B_p, \sigma^p(b_0), \ldots, \sigma^p(b_{h-1}), B'_p]$ est un p-découpage naturel du mot $B_p\sigma^p(b_0)\ldots\sigma^p(b_{h-1})B'_p$ au rang $j+|A_p|$.

De plus au rang j le mot A_pB_p apparaît comme suffixe d'un $\sigma^p(b_{-l}\ldots b_{-1})$, où $b_{-l}, \ldots, b_{-1}$ sont des lettres telles que $[\sigma^p(b_{-l}), \ldots, \sigma^p(b_{-1})]$ est un p-découpage naturel du mot $\sigma^p(b_{-l})\ldots\sigma^p(b_{-1})$ au rang $j+|A_pB_p|-|\sigma^p(b_{-l}\ldots b_{-1})|$.

Nous choisirons pour l le plus petit entier naturel tel que $|\sigma^p(b_{-l}\ldots b_{-1})|>|\sigma^p(a_{-1})B_p|$, quitte, s'il n'y a pas assez de lettres à gauche de u_j, à remplacer j par un rang plus grand, où les découpages seront les mêmes, ce qui est toujours possible par minimalité.

Comme σ est primitive, h est majoré en fonction de k indépendamment de p, et pour une infinité de valeurs de p, ce sont les mêmes mots $a_{-1}\ldots a_{k-1}$ et $b_{-l}\ldots b_{h-1}$ qui interviennent, étant en nombre fini.

(3) Si pour deux entiers p et q $(p<q)$ ainsi choisis, on a:

$$\sigma^p(a_0\ldots a_{k-1})=B_p\sigma^p(b_0\ldots b_{h-1})B'_p \quad \text{et}$$

$$\sigma^q(a_0\ldots a_{k-1})=B_q\sigma^q(b_0\ldots b_{h-1})B'_q,$$

avec $B_q\neq\sigma^{q-p}(B_p)$, alors

$$\sigma^q(a_0\ldots a_{k-1})=B_q\sigma^q(b_0\ldots b_{h-1})B'_q=\sigma^{q-p}(B_p)\sigma^q(b_0\ldots b_{h-1})\sigma^{q-p}(B'_p),$$

ce qui entraine que $\sigma^q(a_0\ldots a_{k-1})$ commence par une puissance d'un mot v d'autant plus grande que k, fixé au départ, est grand.

Si cette situation peut être réalisée pour des valeurs croissantes de k, on déduit de la Section 2 que u est périodique, ce que nous excluons ici.

(4) Si, pour presque toutes les valeurs de k, pour tous les rangs correspondant à $a_{-1},\ldots,a_{k-1}$ et à $b_{-l},\ldots,b_{h-1}$, les mots B_p obtenus sont tels que si $p<q$, $B_q=\sigma^{q-p}(B_p)$, nous choisissons pour k une des valeurs qui conviennent.

Le mot R_q apparaît:

- à un rang $i=i_{(q)}$ appartenant à E_1 avec le q-découpage naturel $[A_q,\sigma^q(a_0),\ldots,\sigma^q(a_{k-1})]$,
- à un rang $j=j_{(q)}$ n'appartenant pas à E_1 sous la forme $A_q\sigma^{q-p}(B_p)\sigma^q(b_0\ldots b_{h-1})B'_q$, le q-découpage $[\sigma^q(b_{-l}),\ldots,\sigma^q(b_{h-1}),B'_q]$ étant naturel au rang $j_{(q)}+|A_qB_q|-|\sigma^q(b_{-l}\ldots b_{-1})|$.

D'autre part on trouve au rang $j_{(p)}$ le mot R_p sous la forme $A_pB_p\sigma^p(b_0\ldots b_{h-1})B'_p$, le p-découpage $[\sigma^p(b_{-l}),\ldots,\sigma^p(b_{h-1}),B'_p]$ étant naturel au rang $j_{(p)}+|A_pB_p|-|\sigma^p(b_{-l}\ldots b_{-1})|$. On déduit alors du fait que $|\sigma^q(b_{-l}\ldots b_{-1})|>|\sigma^q(a_{-1})B_q|$, en appliquant σ^{q-p} et par comparaison avec le rang $i_{(q)}$ que le mot R_q apparaît à un rang i' appartenant bien à E_1 sous la forme $A_q\sigma^{q-p}(B_p)\sigma^q(b_0\ldots b_{h-1})B'_q$, le q-découpage $[\sigma^q(b_{-l}),\ldots,\sigma^q(b_{h-1})]$ étant naturel au rang $j_{(q)}+|A_qB_q|-|\sigma^q(b_{-l}\ldots b_{-1})|$ et le $(q-p)$-découpage du mot $\sigma^q(a_{-1})\sigma^{q-p}(B_p)$ étant le même au rang $i_{(q)}+|A_q|-|\sigma^q(a_{-1})|$ et au rang $i'+|A_q|-|\sigma^q(a_{-1})|$. Ainsi les entiers $j'=j_{(q)}+|A_q\sigma^{q-p}(B_p)|$ et $i''=i'+|A_q\sigma^{q-p}(B_p)|$ appartiennent à E_q et on peut leur appliquer la remarque préliminaire: à gauche de ces deux rangs on trouve dans u le même mot avec le même 1-découpage, d'autant plus long que q est grand. Or i' appartient à E_1, mais pas $j_{(q)}$. On en déduit qu'on est dans la situation de l'énoncé du théorème.

La réciproque est immédiate. □

Démonstration du Théorème 3.1bis. Les idées directrices de cette démonstration sont celles de la démonstration du Théorème 3.1, en faisant simultanément cette fois le même travail à gauche et à droite des rangs considérés.

Supposons que σ n'est pas bilatéralement reconnaissable.

Comme u n'est pas périodique, nous montrons comme dans la démonstration précédente (utilisation des entiers i'' et j' et de leurs homologues à gauche) qu'il existe un ensemble infini I d'entiers naturels, des lettres $c_0 \ldots c_{r-1}$ et des lettres $d_0 \ldots d_{s-1}$ telles que pour tout élément q de I il existe des entiers $m_{(q)}$, $n_{(q)}$, $m'_{(q)}$ et $n'_{(q)}$ appartenant à E_q tels que:

- $C_q = u[m_{(q)}, n_{(q)}] = u[m'_{(q)}, n'_{(q)}]$,
  le q-découpage naturel de C_q est $[\sigma^q(c_0), \ldots, \sigma^q(c_{r-1})]$ au rang $m_{(q)}$,
- le q-découpage naturel de C_q est $[\sigma^q(d_0), \ldots, \sigma^q(d_{s-1})]$ au rang $m'_{(q)}$,
- le 1-découpage de C_q n'est pas le même au rang $m_{(q)}$ et au rang $m'_{(q)}$.

Mais si on est dans cette situation au rang q, pour tout entier $p > q$ appartenant à I, le 1-découpage de C_p doit être le même le même au rang $m_{(p)}$ et au rang $m'_{(p)}$, ce qui est contradictoire, d'où le résultat. □

Corollaire 3.2. *Toute substitution primitive de longueur constante (i.e. telle que pour tous a et b dans $\mathcal{A}$, $|\sigma(a)| = |\sigma(b)|$) et admettant un point fixe non périodique est reconnaissable.*

Remarques et exemples. (1) La substitution définie sur $\mathcal{A} = \{a, b, c, d\}$ par $\sigma(a) = ad$, $\sigma(b) = a$, $\sigma(c) = d$ et $\sigma(d) = bc$ est reconnaissable: $\sigma(c)$ est suffixe strict de $\sigma(a)$, mais a est toujours suivi par d, et c par a dans le point fixe u.

(2) Si pour toutes lettres distinctes a et b de $\mathcal{A}$, $\sigma(a)$ est suffixe strict de $\sigma(b)$ ou inversement, et si u n'est pas périodique, σ n'est pas reconnaissable: en effet il existe une suite strictement croissante d'entiers naturels $(L_n)_{n \in \mathbb{N}}$ telle que pour tout n il existe un mot B de longueur L_n prolongeable à gauche de deux façons différentes en éléments de L_u.

Par exemple la substitution définie sur $\mathcal{A} = \{a, b\}$ par $\sigma(a) = aba$ et $\sigma(b) = ba$ n'est pas reconnaissable.

(3) Si σ n'est pas reconnaissable, aucun itéré de σ n'est reconnaissable.

(4) Si σ est reconnaissable, et injective sur les lettres, les itérés de σ sont reconnaissables.

Références

[1] F.M. Dekking, Combinatorial and statistical properties of sequences generated by substitutions, Thèse, Nijmegen, 1980.

[2] B. Host, Valeurs propres des systèmes dynamiques définis par des substitutions de longueur variable, *Ergodic Theory and Dynamical Systems* **6** (1986) 529–540.

[3] M. Lothaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983).

[4] J.C. Martin, Minimal flows arising from substitutions of non constant length, *Math. Systems Theory* **7** (1973) 73–82.

[5] M. Queffélec, Substitution dynamical systems – Spectral analysis, Lecture Notes in Mathematics, Vol. 1294 (Springer, Berlin, 1987).

[6] P. Séébold, Propriété combinatoires des mots infinis engendrés par certains morphismes, Thèse de Doctorat, 1985, Rapport L.I.T.P. 85-16.

[7] P. Séébold, An effective solution to the DOL periodicity problem in the binary case, *EATCS Bull.* **36** (1988) 137–151.

[8] A. Thue, Über unendliche Zeichenreihen (1906), Selected Mathematical Papers of Axel Thue (Universitetsforlaget, 1977).